

모런 고유벡터 공간 필터링 기반 공간 가변 계수를 이용한 서울시 아파트 가격 결정요인의 공간적 이질성 탐색

구형모*

Exploring Spatial Heterogeneity in Factors for Apartment Prices in Seoul Using Moran Eigenvector Spatial Filtering Based on Spatially Varying Coefficients

Hyeongmo Koo*

요약 : 공간 가변 계수 모형은 회귀 분석 기반의 공간적 이질성 탐색을 위한 도구로 활용되어 왔다. 그 중 하나인 공간 가중 회귀 모형(GWR)은 모형의 설명력 부족과 계수 간의 다중공선성과 같은 한계에도 불구하고 대다수의 주택 가격 결정요인의 공간적 이질성 탐색 연구에서 널리 이용되었다. 본 연구에서는 GWR 모형의 한계 극복을 위한 대안으로 모런 고유벡터 공간 필터링(MESF) 기반 모형을 서울시 아파트 매매가격의 결정요인을 탐색하기 위하여 적용하고 그 결과를 GWR, 그리고 전역적 MESF 모형과 비교하였다. 본 연구 결과에서는 모형의 설명력, 잔차의 공간적 자기상관 유무, 그리고 추정된 공간 가변 계수 간의 다중공선성 유무를 기준으로 MESF가 GWR보다 향상된 결과를 보였다. 또한 국지적으로 매우 다른 양상을 보이는 각 결정요인들의 아파트 매매가격 대한 효과는 전역적 수준에서의 결정요인의 효과가 하위지역에서 동일하지 않는 점을 보여주며 공간적 이질성 탐색이 필수적임을 시사한다.

주요어 : 아파트 매매가격, 공간 가변 계수, 모런 고유벡터 공간 필터링, 공간 가중 회귀

Abstract : Spatially varying coefficients models have been used as a tool for exploring spatial heterogeneity in regression coefficients. But, among various spatially varying coefficients models, only a geographically weighted regression (GWR) has been used for exploring spatial heterogeneity in factors for housing prices although it suffers from its low explanatory power and high multicollinearity between local coefficients. This study applies another type of a spatially varying coefficients model based on Moran eigenvector spatial filtering (MESF) to explore factors to apartment sale prices. The analysis results are compared to those of GWR and global MESF. The result shows MESF outperforms GWR in terms of explanatory power, spatial autocorrelation in model residuals, and multicollinearity between local coefficients. In additions, a variation in local coefficients against their global coefficients suggests that exploring spatial heterogeneity is necessary for a housing price estimation model.

Key Words : Apartment sale price, Spatially varying coefficients, Moran eigenvector spatial filtering, Geographically weighted regression

*중국 난징사범대학교 지리과학학원 박사후 연구원(Post-doctoral Researcher, School of Geography, Nanjing Normal University, China, hmo.koo@gmail.com)

I. 연구배경 및 목적

주택은 인간의 생존을 위해 필수적 요소 중 하나로 그 가격은 주택의 구매자가 되는 개인의 삶 뿐 아니라, 주택이 위치한 지역의 전·출입과 과세에도 매우 큰 영향을 미친다. 따라서 주택 가격에 대한 정확한 예측은 지역 계획 및 운영에 필수적이며(진찬우·이건학, 2014), 이를 위하여 주택 가격에 영향을 미치는 주택의 물리적, 지역적, 그리고 입지적 특성과 같은 다양한 가격 결정요인들에 대한 탐색이 선형 회귀 모형과 공간 회귀 모형 등을 이용하여 활발히 이루어져 왔다(전경구, 2012; 김소연·김영호, 2013; 김민성·박세운, 2014; 임준호·홍성호, 2015). 하지만 각 지역별로 주택의 소유와 공급 현황 나아가 지역적 특성이 다르기 때문에 주택 가격의 결정요인들의 전역적 수준의 영향이 하위 지역에서 역시 동일하게 나타나지는 않는다(김갑영, 2017), 따라서 전역적 수준에서 추정된 주택 가격 회귀 모형은 주택 가격 결정요인들의 국지적 차이를 설명하기에는 제한적이다(진찬우·이건학, 2014).

공간 가변 계수(spatially varying coefficients)는 회귀 분석 내 각 요인들의 전역적 계수의 공간적 이질성을 표현하기 위해서 활용되어 왔으며(Helbich and Griffith, 2016), 공간 확장 모형(spatial expansion method)(Casetti, 1972; 1997), 이동 창 회귀 모형(moving window regression)과 공간 가중 회귀(geographically weighted regression; 이하 GWR)(Fotheringham *et al.*, 2002), 그리고 모란 고유벡터 필터링(Moran eigenvector spatial filtering; 이하 MESF) 기반 모형(Griffith, 2008)이 공간 가변 계수 추정을 위한 대표적 모형이다. 국내의 주택 가격 결정요인의 공간적 이질성을 탐색하기 위한 연구에서는 GWR을 주로 이용하여 왔으며(강창덕, 2010; 김혜영·전철민, 2012; 남형권·서원식, 2017; 김정희, 2018), 또한 그 결과를 토대로 주택 가격 결정요인들의 공간적 이질성 탐색의 중요성을 역설하였다. 하지만 GWR의 경우 공간적 이질성 탐색을 위해 주변에 관측지만을 고려하여 국지적 모형을 추정하기 때문에, 모형 추정에 이용되는 관측지의 수가 적어 모델 설명력이 제한적이며 또한 추정된 국지적 회귀 모형의 각 변수들의 계수 간의 다중공선성(multicollinearity)이 발생할 수 있다(Wheeler and Tiefelsdorf, 2005; Griffith, 2008). GWR에 축소추정법(least absolute

shrinkage and selection operator; LASSO)이나 능형(ridge) 회귀 모형을 결합하여 다중공선성은 어느 정도 해소 가능하나(Wheeler, 2009; 진찬우·이건학, 2014; 전창우 등, 2018), 제한된 관측지 수로 인한 GWR의 모델 설명력의 부족은 여전히 문제로 남아 있다(Helbich and Griffith, 2016). 상대적으로 가장 최근에 개발된 MESF 기반의 공간 가변 계수 모형(Griffith, 2008)은 다른 공간 가변 계수 모형보다 대체로 높은 모형 설명력을 지니며 공간 가변 계수들 간의 다중공선성의 요인을 공통 고유벡터를 이용하여 쉽게 추적할 수 있다는 장점이 있다고 알려져 있지만, 모형 적용의 기술적 어려움으로 인하여 국내 및 해외의 연구에서 제한적으로 다루어져 왔다(Chun and Griffith, 2013; Helbich and Griffith, 2016).

본 연구의 목적은 공간 가변 계수를 이용하여 주택 가격을 결정하는 요인들에 대한 공간적 이질성을 탐색함과 동시에 MESF 기반 모형의 국내 연구에의 적용가능성을 살펴보는 것이다. 본 연구에서는 국내에서 평균 주택 가격이 가장 높은 서울, 그 중에서도 서울시 내 대부분의 주택을 차지하고 있는 유형인 아파트의 매매가격에 초점을 맞추어 그 가격 결정요인을 탐색하고자 한다. 이를 위하여 본 연구는 다음과 같은 순서로 이루어진다. 먼저, MESF 기반 음이항(negative binomial) 회귀 모형을 이용하여 서울시 내 아파트 매매가격 결정요인들의 전역적 계수를 살펴본다. 그 후 아파트 매매가격을 결정하는 요인들의 공간적 이질성을 살펴보기 위하여 공간 가변 계수를 추정한다. 공간 가변 계수는 각각 GWR과 MESF를 기반으로 추정한 후, 모형의 평가를 위하여 두 모형 결과의 추정된 계수 간의 다중공선성, 모형의 설명력, 그리고 모형 잔차의 공간적 자기상관 유무를 비교한다. 마지막으로 공간 가변 계수의 공간적 분포를 통해 서울시 아파트 매매가격 결정요인의 공간적 이질성을 탐색한다. 본 연구에서 MESF 기반 공간 가변 계수 모형은 SAAR(spatial analysis using ArcGIS engine and R)(Koo *et al.*, 2018) 소프트웨어를 이용하여 추정하였으며,¹⁾ 해당 소프트웨어는 그래픽 사용자 인터페이스(graphic user interface: GUI)를 제공하여 일반화 선형 모형(generalized linear model), 그리고 공간 오차 모형이나 MESF 등과 같은 공간 회귀 모형을 쉽게 적용 가능하다는 장점이 있다. 그리고 GWR은 일반화 선형 모형을 적용한 확장이 필요하기 때문에 통계 패키지 R을 이용하여 추정하였다.

II. 연구방법

1. 음이항 회귀 모형과 변수 선정

우선 본 분석에서는 음이항 회귀 모형을 이용하여 아파트 매매가격을 분석하였다. 일반적으로 아파트의 매매가격과 같이 최솟값이 정해져 있으나 최댓값이 정해지지 않은 변수를 분석할 경우에는 포아송 회귀 분석을 사용하는 것이 합당하다(Griffith and Haining, 2006). 하지만 본 분석에서 포아송 회귀 모형을 사용할 경우 매우 높은 편차 통계(deviance statistic) 값이 관측되었고, 이는 평균과 분산이 매우 다른 과산포(overdispersion)가 존재한다는 것을 의미한다. 이러한 과산포는 포아송 분포의 기본 가정인 평균과 분산이 같다는 것을 충족하지 못할 뿐 아니라 표준 오차(standard error) 계산에도 심각한 오류를 발생시켜 계수의 유의성 검사 신뢰도의 하락을 가져온다(Chun and Griffith, 2013:77). 따라서 본 분석에서는 음이항 회귀 모형을 사용하였고 그 결과 적합한 편차 통계 값을 얻을 수 있었다.²⁾

회귀 모형에 사용된 변수는 다음과 같다(표 1, 그림 1). 주택시장은 지리적으로 이질적인 현상이며 그 범위는 행

정구역과 일치하지 않을 수 있기 때문에(김감영, 2016; 조대현, 2018) 본 분석에서는 모든 변수들은 500m×500m 크기의 격자 단위로 요약하여 이용하였다. 따라서 아파트 매매자료가 존재하지 않는 격자를 제외한 전체 1,481개의 격자로 연구지역을 표현하였고, 아파트가 존재하지 않는 격자는 전 지도에 걸쳐 회색으로 표시하였다(그림 1, 3, 4). 종속변수로는 국토교통부에서 제공하는 아파트의 매매 실거래 자료를 이용하였으며³⁾ 시간 범위는 2017년 1월 1일부터 2018년 12월 31일까지의 2년간 매매자료이다. 일반적으로 아파트 매매가격은 아파트의 크기와 밀접한 관계를 지니고 있으므로(전경구, 2012; 김민성·박세운, 2014; 임준호·홍성효, 2015; 강수진·서원석, 2016) 본 분석에서는 아파트 크기를 음이항 회귀 모형의 오프셋(offset)으로 사용하였다. 그 결과 실제 모형은 단위면적당 아파트 매매가격(그림 1의 (가))을 종속변수로 이용한 것과 같은 효과를 지닌다. 상세 자료 처리 방법은 먼저, 거래된 아파트의 주소를 기반으로 지오코딩을 실시한 후, 각 거래의 아파트 매매가 이용하여 격자 내의 전체 매매가의 평균가격을 계산하고 이를 종속변수로 이용하였다. 마찬가지로 아파트의 크기도 각 격자 단위로 평균값을 구한 뒤 회귀 모형의 오프셋으로 사용하였다.

표 1. 분석에 활용된 변수

구분	변수명	단위	설명	출처
종속변수	아파트 매매가	만원	격자 기반 아파트 실거래 매매가격의 평균(2017-2018)	국토교통부 아파트 매매실거래
오프셋	아파트 크기	m ²	격자 기반 거래된 아파트 크기의 평균(2017-2018)	국토교통부 아파트 매매실거래
독립변수	노후 연수	년	격자 기반 거래된 아파트 건축 연도의 평균으로부터 계산(2018년 기준)	국토교통부 아파트 매매실거래
	보육시설 접근성	km	격자 중심점에서 가장 가까운 보육시설(어린이집·유치원)까지의 도로상 이동거리	국토교통부 생활기반시설 접근성
	공원시설 접근성	km	격자 중심점에서 가장 가까운 도시공원까지의 도로상 이동거리	국토교통부 생활기반시설 접근성
	지하철역 접근성	km	격자 중심점에서 가장 가까운 지하철역까지의 도로상 이동거리	서울교통공사 지하철역 위치·국가교통데이터베이스
	방과후 프로그램 수	수/100명	초등학교의 학생 수 100명 당 방과후 프로그램 수	학교알리미
	특목·자립고 진학비율	%	중학교의 전체 졸업생 대비 특목목적고 및 자립형 고교의 진학생 비율	학교알리미
	대학 진학률	%	고등학교 전체 졸업생 대비 대학 진학자 수의 비율	학교알리미

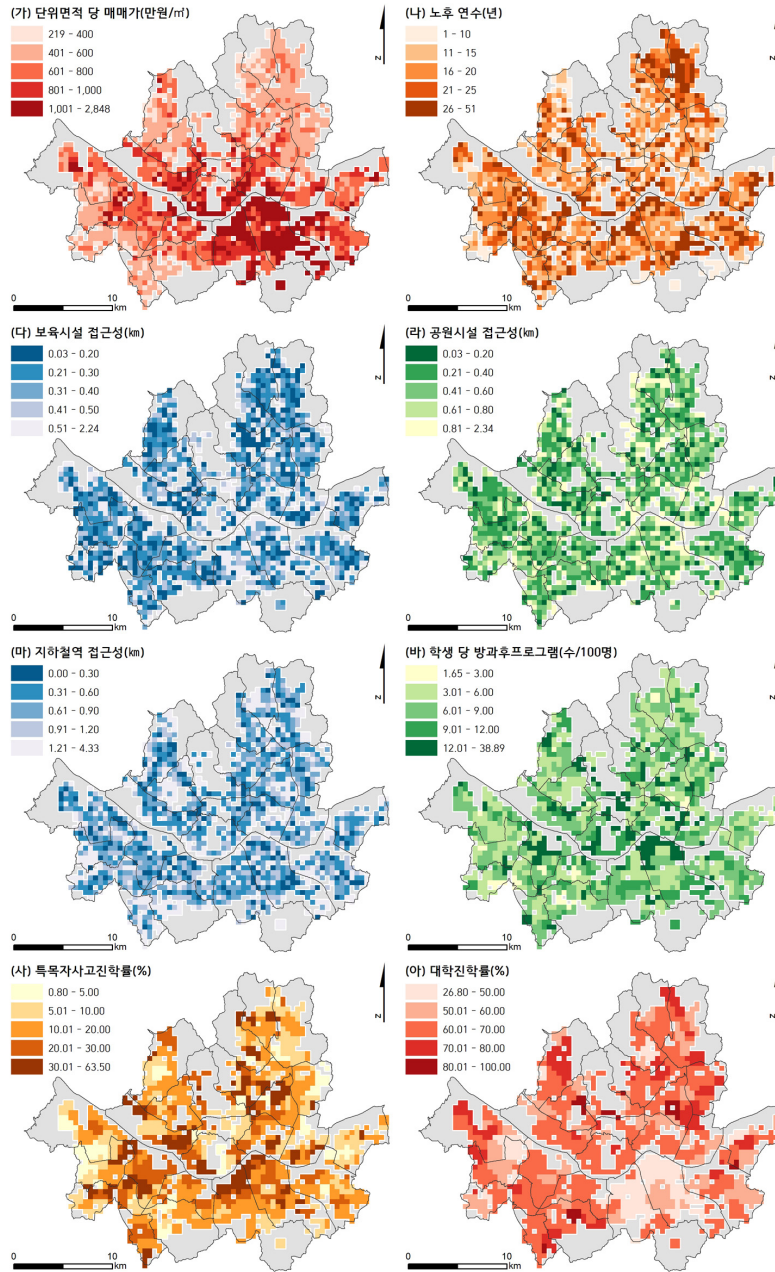


그림 1. 아파트 단위면적 당 매매가격과 선택된 가격 결정요인

주택의 가격은 크게 주택의 입지 특성, 물리적 특성, 주택 시장의 속성, 그리고 주택의 시·공간적 종속성에 따라 결정된다(김소연·김영호, 2013). 본 분석에서는 주택의 물리적 특성과 입지 특성과 관련된 변수들을 주된 아파트 가격 결정요인으로 사용하였다. 먼저, 주택의

물리적 특성은 아파트의 노후 연수를 이용하여 표현하였다. 다음으로 주택의 입지 특성 중 주거환경은 공원시설의 접근성, 사회 및 교통 편의성은 보육시설의 접근성과 지하철역까지의 접근성, 마지막으로 교육환경은 초등학교의 학생 100명 당 방과후 프로그램 수, 중학교의

특수목적고 및 자립형 고교의 진학률, 고등학교의 대학 진학률을 사용하였다(표 1, 그림 1). 먼저 아파트의 노후 연수(그림 1의 (다))는 기존 연구들(정수연, 2006; 박운선·임병준, 2010; 전경구, 2012; 김민성·박세운, 2014; 강수진·서원석, 2016)에서 아파트 매매가격과 음의 상관관계가 관찰되었다. 본 분석에서 아파트 노후 연수는 국토교통부의 매매실거래 자료에서 거래된 아파트의 건축연도를 기반으로 각 격자 단위로 그 평균값을 계산한 뒤 2019년을 기준으로 각 격자별로 획득하였다.

다음으로 보육시설과 공원시설의 접근성을 아파트 매매가격의 결정요인으로 고려하였다(그림 1의 (다), (라)). 기존 연구에서는 보육시설(김소연·김영호, 2013; 홍하연·이주형, 2015)과 공원시설(반운선·임병준, 2010; 전경구, 2012)까지의 접근성이 감소할수록 아파트 매매가격은 하락하는 경향을 보였다. 본 연구에서 보육시설의 접근성과 공원시설의 접근성은 국토교통부의 생활기반시설 접근성 자료를 이용하였다.⁴⁾ 이 중 보육시설 접근성은 2016년 4월을 기준으로 교육부 유아교육 정책과 전국유치원과 2017년 7월 기준 보건복지부 보육정책과 어린이집 현황 자료를 기반으로 각 격자의 중심으로 부터 가장 가까운 유치원이나 어린이집의 도로이동거리(km)를 국가교통 데이터베이스를 이용하여 추정한 것이다. 공원시설의 접근성은 동일한 방법으로 가장 가까운 도시공원⁵⁾까지의 도로이동거리를 나타낸 것이다. 공원의 위치는 한국토지주택공사 공간정보처의 2016년 자료를 기반으로 추정되었다. 따라서 도로이동거리의 증가는 접근성의 감소를 의미하며, 본 분석에서는 이 두 변수와 아파트 매매가격 간에는 음의 상관관계가 예상된다.

다음은 지하철역까지의 접근성이다(그림 1의 (마)). 지하철역의 접근성은 기존의 연구들에서도 자주 활용된 결정요인으로 해당 연구들에서 아파트의 매매가격에 양의 상관관계를 보이는 것으로 분석되었다(정수연, 2006; 박운선·임병준, 2010; 전경구, 2012; 김소연·김영호, 2013; 김민성·박세운, 2014; 임준홍·홍성호, 2015; 홍하연·이주형, 2015; 강수진·서원석, 2016). 다시 말해, 지하철역 접근성이 증가할수록, 혹은 지하철역까지의 거리가 감소할수록, 아파트 매매가격은 대체로 증가하는 경향을 보였다. 본 분석에서 지하철역의 접근성 역시 앞의 두 접근성 지표와 마찬가지로 각 격자 중심으로부터 가장 가까운 지하철역까지의 도로이동거리로 표현하였다. 지하철역의 위치는 서울교통공사 2017년 9월 23일

기준 수도권 지하철역 공간좌표를 기반으로 신설된 지하철역(예: 9호선 중앙보훈병원역)을 추가하여 나타내었고 도로이동거리 계산을 위해서는 국가교통데이터베이스(Korea Transport Database; KTDB)를 이용하였다.⁶⁾

마지막으로 교육요인이다(그림 1의 (바), (사), (야)). 국내의 사례에서 교육요인은 아파트의 매매가격을 결정하는 중요한 요인으로 인식되어져 왔으며, 이를 분석에 활용하기 위하여 학군 및 학업 성취 등 다양한 변수가 사용되어 왔다(김경민·이양원, 2007; 전경구, 2012; 김소연·김영호, 2013; 임준홍·홍성호, 2015; 홍하연·이주형, 2015). 본 분석에서는 초등학교, 중학교, 그리고 고등학교 각각의 교육 환경 및 성취가 아파트 매매가격에 미치는 영향을 분리하여 확인하고자 하였다. 따라서 초등학교에서는 학생 수 100명 당 교과 및 특기적성을 모두 포함한 방과후 프로그램의 수를, 중학교에서는 일반 중학교의 특수목적고 및 자율고 진학비율(이하 특목·자립고 진학비율)을, 그리고 고등학교에서는 일반 고등학교의 대학 진학률을 사용하여 교육 환경 및 성취를 나타내고자 하였다. 모든 자료는 한국교육학술정보원에서 총괄·관리하는 학교알리미⁷⁾에서 획득하였다. 획득된 자료를 해당 학교의 주소정보와 결합하고, 지도 코딩을 통해 해당 학교를 좌표화하였다. 마지막으로 각 격자의 중심점에서부터 직선거리로 가장 가까운 학교의 정보를 이용, 각 격자에 해당 변수 값을 할당하였다.

2. 공간 가변 계수 모형(spatially varying coefficients models)

본 분석에서 공간 가변 계수는 MESF(Griffith, 2008)와 GWR(Fotheringham *et al.*, 2002)을 이용하여 추정하였다. 먼저, 전역적 계수 추정을 위한 MESF는 공간자료에 존재하고 있는 공간적 자기상관이 야기할 수 있는 회귀 분석의 잔차 독립성의 위배를 극복하기 위하여 개발되었다(Griffith, 2003).

$$(\mathbf{I} - \mathbf{11}^T/n)\mathbf{C}(\mathbf{I} - \mathbf{11}^T/n) \quad (1)$$

간략히 MESF는 식 (1)에서 추출된 고유벡터들을 독립변수로 활용하여 회귀 분석 모형 내에서 공간적 자기상관을 고려한다. 여기서 $\mathbf{C}$ 는 이항 공간 가중 행렬(binary spatial weights matrix), $\mathbf{I}$ 는 항등 행렬, $\mathbf{1}$ 은 1로 이루어

진 벡터, 그리고 τ 는 벡터의 전치를 의미한다. 추출된 고유벡터들은 서로 직교하는 성질을 가지고 있을 뿐 아니라 해당 연구지역이 가질 수 있는 모든 공간적 자기상관의 형태를 표현할 수 있기 때문에(Tiefelsdorf and Boots, 1995; 이상일 등, 2015; 2016), 회귀 분석 내의 독립변수로 사용이 가능하다. 또한 MESF는 선형 회귀 모델뿐 아니라 일반화 선형 모델에도 쉽게 적용 가능하다는 장점이 있다(Griffith, 2002; 구형모, 2018). 본 분석에서 이항 공간 가중 행렬은 퀸(queen) 방식으로 생성하였고, 한강이나 산 등 자연환경으로 인해 단절된 격자의 경우 실제 도로를 기반으로 주변 지역과 인접할 수 있도록 공간 자료를 수정하여 가능한 분리된 격자가 없게 하였다. 마지막으로 아파트 매매가는 강한 양의 공간적 자기상관⁸⁾을 보이기 때문에, 전체의 고유벡터 중 첫 번째 고유치에 비해 50% 이상의 고유치를 가지고 있는 고유벡터들을 고유벡터 후보군으로 선정하였다. 회귀 모형 잔차의 공간적 자기상관 확인을 위해서 잔차는 피어슨 잔차(Pearson residuals)를 이용하여 추출하였고, 이를 바탕으로 전역적 모런 통계량(Moran's I)을 계산한 후 순열 검정법(permutation test)을 이용하였다(Lin and Zhang, 2007).

MESF는 회귀 분석에서 공간적 이질성을 탐색할 수 있는 공간 가변 계수를 추정할 수 있도록 다음과 같이 확장 가능하다 (Griffith, 2008; Helbich and Griffith, 2016).

$$Y = \left(\beta_0 \mathbf{1} + \sum_{k_0=1}^{K_0} \mathbf{E}_{k_0} \beta_{k_0} \right) + \sum_{p=1}^P \left(\beta_p \mathbf{1} + \sum_{k_p=1}^{K_p} \mathbf{E}_{k_p} \beta_{k_p} \right) \cdot X_p + \epsilon \quad (2)$$

식 (2)에서 Y 는 종속변수의 값, 본 분석에서는 아파트의 매매가격이며, X_p 는 각각의 독립변수(p)의 값, $\mathbf{E}_{k_p}$ 는 종속변수(p)를 설명하기 위하여 선택된 고유벡터, P 는 전체 독립변수의 수, K_p 는 전체 고유벡터 후보군의 수, β 는 각각의 벡터들에 대해 추정된 계수, 그리고 $\cdot$ 는 각 행렬 요소의 곱을 의미한다. 식 (2)의 앞부분은 공간 가변 회귀 상수를, 그리고 뒷부분은 공간 가변 계수를 나타낸다. 식 (2)는 식 (3)과 같이 재정렬이 가능한데,

$$Y = \left(\beta_0 \mathbf{1} + \sum_{p=1}^P X_p \cdot \beta_p \right) + \left(\sum_{k=1}^K \mathbf{E}_k \beta_k + \sum_{p=1}^P \sum_{k=1}^K X_p \cdot \mathbf{E}_k \beta_k \right) + \epsilon \quad (3)$$

식 (3)에서의 앞부분은 전역적인 계수, 그리고 뒷부분은

그 전역적 계수의 국지적 변화를 표현한다. 위 식을 통해 MESF의 두 가지 장점을 알 수 있다(Griffith, 2008; Helbich and Griffith, 2016). 먼저, MESF 기반으로 추정된 공간 가변 계수는 전역적 계수를 중심으로 국지적으로 변화하는 특징을 보인다. 따라서 본 분석에서는 공간적 가변 계수를 전역적 계수 값을 기준으로 지도화하고 그 공간적 이질성을 탐색할 것이다. 다음으로 공간 가변 계수 간의 다중공선성이 존재할 경우 각 변수 간 공통의 고유벡터를 기반으로 쉽게 설명 가능하다는 점이 있다. 동시에 본 분석과 같이 분석에 많은 수의 관측치를 이용할 경우(전체 1,481개의 격자) 공통의 고유벡터를 가질 확률이 적어지기 때문에, 공간 가변 계수 간의 다중공선성은 감소할 것으로 예상된다.

다음으로 GWR은 국내 연구에서 회귀 모형의 공간적 이질성을 탐색하기 위해 사용되는 대표적인 방법론이다(김익희, 2005; 전창우 등, 2018). 간략히 GWR은 전역적인 하나의 회귀 모형을 이용하는 것이 아니라 각 공간적 위치를 중심으로 거리 조락에 따른 가중치를 부여하여 그 위치별로 국지적 회귀 모형을 추정하는 형태를 지닌다(Fotheringham *et al.*, 2002). 다시 말해, GWR은 탐색반경(bandwidth)을 설정하여 그 범위에 속하는 관측치들에 대해 거리조락의 가중치를 이용하여 회귀 분석에 이용하고, 범위 밖의 관측치는 회귀 분석의 대상으로 고려하지 않는다. 따라서 탐색반경과 거리조락의 표현 방법 설정이 GWR 추정의 두 가지 필수적 요소이다. 먼저, 일반적으로 최적의 탐색반경은 교차검증(cross-validation)을 이용하여 추정되는데(Fotheringham *et al.*, 2002), 본 분석에서는 GWR과 MESF 기반의 공간 가변 계수 모형의 직접적인 비교를 위하여 가장 높은 설명력(pseudo- R^2) 값을 보이는 탐색반경을 이용하여 GWR을 추정하였다. 상세히 설명하자면 탐색반경의 범위 500m에서 5,000m 값 사이에서 가장 높은 모델 설명력을 보이는 탐색반경을 황금 분할 탐색(golden section search with successive parabolic interpolation) 기법을 이용하여 추정하였다(Brent, 1973). 다음의 거리 조락에 따른 가중치는 GWR에서 일반적으로 널리 사용되는 가우시안 커널(Gaussian kernel)을 사용하였다(Redfeam, 2009; Helbich and Griffith, 2016). GWR 역시 음이항 회귀 모형을 기반으로 동일한 종속변수와 독립변수를 이용하여 추정되었다. 이를 위하여 R의 sqgwr 패키지⁹⁾의 ggwr함수를 음이항 회귀 모형을 기반으로 수정하여 사용하였다.

III. 연구 결과

1. MESF와 GWR 기반의 공간 가변 계수 모형 결과 비교

공간 가변 계수 모형을 추정하기에 앞서 전역적 MESF 모형의 분석 결과는 다음과 같다(표 2). 전역적 MESF 모형 추정을 위해서 단계적 회귀 분석 기법(stepwise regression)을 이용하여 고유벡터를 선택하였다. 고유벡터 후보군은 전체 1,481개의 고유벡터 중 167개의 고유벡터이며, 그 중 95개의 고유벡터가 아파트 매매가격을 설명하기 위하여 최종적으로 선택되었다. 그 결과 전역적 MESF 모형의 설명력(pseudo- R^2)은 0.765로 상대적으로 높은 값을 보이나, 모형의 잔차에서는 유의미한 공간적 자기상관이 확인되었다.¹⁰⁾ 아파트 매매가와 신뢰도 99%에서 유의미한 상관관계를 보이는 변수는 아파트의 노후 연수, 특목·자립고 진학비율, 그리고 대학 진학률이다. 그 중 아파트 노후 연수와 특목·자립고 진학비율은 예상과 같이 각각 양과 음의 상관관계를 보인다. 즉 아파트가 최근에 지어지고, 주변에 특목·자립고 진학비율의 진학률이 높을 중학교가 있을 경우 아파트 매매가는 상승하는 경향을 보인다는 것을 의미한다. 반면 기존의 연구와는 다르게 대학 진학률은 아파트 매매가격과 음의 상관관계를 보이는데, 이는 아파트 주변에 대학 진학률이 높은 고등학교가 있을 경우 오히려 아파트 매매가격은 하락하는 경향을 보인다는 것이다. 특징적으로 서

초구와 강남구는 다른 지역에 비해 매우 높은 단위면적당 아파트 매매가(그림 1의 (가))를 보이지만, 매우 낮은 대학 진학률(그림 1의 (아))을 보이는 대신 진학과 취업이 아닌 기타 진로에 높은 비율을 보인다. 이는 해당 지역에 거주하는 학생들이 진학할 대학이 본인에 기대에 만족치 않을 경우 재수를 선택하는 등 대학 진학에 전략적으로 선택할 할 수 있는 여건이 있음을 의미하는 것으로(김경근, 2019), 본 분석에서의 대학 진학률과 매매가격 간의 음의 관계의 논리적 이해가 가능하게 한다. 다음으로 지하철역까지의 접근성과 방과후 프로그램 수는 신뢰도 95%에서 각각 유의미한 음과 양의 상관관계를 보인다. 즉, 가장 가까운 지하철역까지의 거리가 감소할수록, 그리고 아파트 주변에 많은 상대적으로 학생수당 높은 방과후 프로그램 수의 비를 가진 초등학교가 위치할 경우 아파트 매매가격은 증가하는 경향을 보인다. 마지막으로 보육시설과 공원시설까지의 접근성은 기존 연구와 같이 아파트 매매가격과 음의 상관관계를 보이거나 유의미하지 않는 것으로 나타났다.

다음은 MESF와 GWR 기반 공간 가변 계수 모형의 결과 비교이다. 상세히 이 두 모형을 모형의 설명력, 모형 잔차의 공간적 자기상관, 그리고 공간 가변 계수의 다중공선성, 이렇게 세 가지 측면에서 비교를 하였다. 각 모형의 개요는 다음과 같다. 먼저 MESF 기반 공간 가변 계수 모형의 경우 단계적 회귀 분석 기법을 통해 167개의 고유벡터 후보군 중 각각 아파트 노후 연수 76개, 보육시설 거리 42개, 공원시설 거리 49개, 지하철역 거리

표 2. 전역적 MESF 모형 요약

변수	계수	표준오차	z-value	p-value
상수항	7.012	0.064	109.498	< 0.01***
노후 연수	-0.004	0.001	-4.377	< 0.01***
보육시설 접근성	-0.017	0.031	-0.538	0.590
공원시설 접근성	-0.025	0.023	-1.110	0.267
지하철역 접근성	-0.023	0.011	-2.188	0.029**
방과후 프로그램 수	0.004	0.002	2.379	0.017**
특목·자립고 진학비율	0.002	0.001	3.664	< 0.01***
대학 진학률	-0.007	0.001	-8.116	< 0.01***
결정계수(Pseudo- R^2)	0.765			
잔차의 모런 통계량(p-value)	0.150 (0.001)			

* 모든 변수의 분산 팽창 인자(Variance Inflation Factor: VIF)는 2보다 작음.

** p-value < 0.05, *** p-value < 0.01.

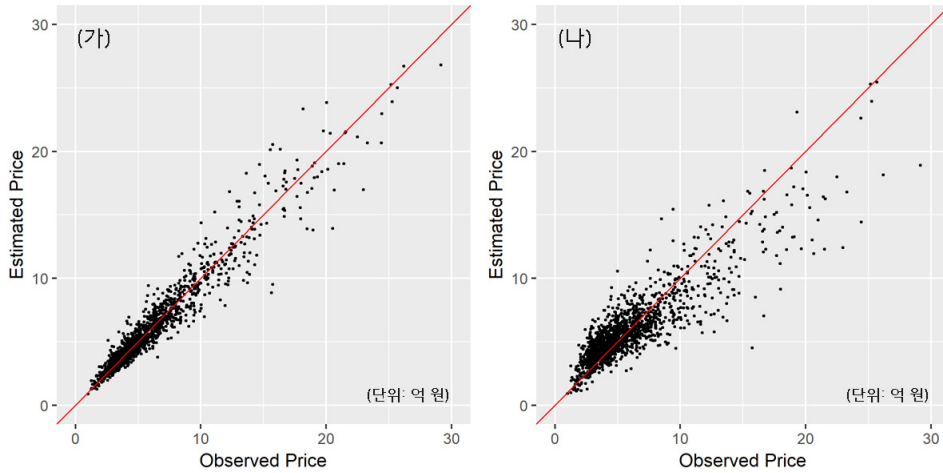


그림 2. 공간 가변 계수 모형의 정확도 비교

* (가) MESF(pseudo-R²=0.932); (나) GWR(pseudo-R²=0.776).

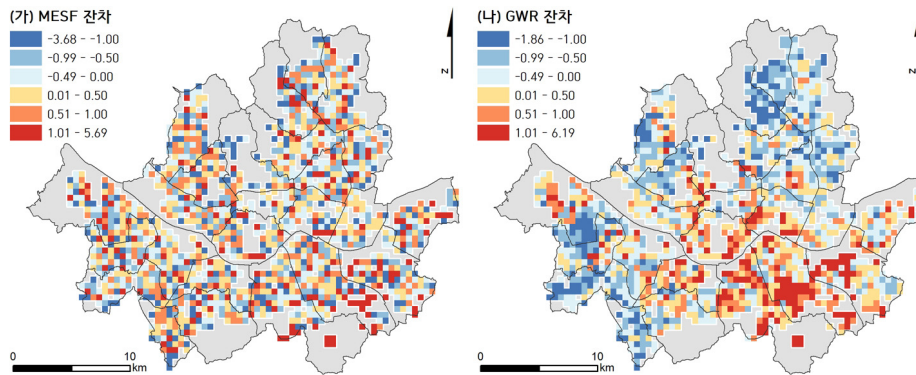


그림 3. 공간가변계수 모형의 잔차 분포

46개, 방과후 프로그램 수 48개, 특목·자립고 진학비율 45개, 그리고 대학 진학률에서 42개의 고유벡터가 전역적 계수의 공간적 이질성을 설명하기 위하여 선택되었다. 음이항 회귀 모형 기반 GWR의 경우 가장 높은 결정 계수(pseudo-R²)을 보이는 탐색반경 1,112m를 가우시안 커널을 이용하여 모형을 구축하였는데, 각 격자의 크기가 500m임을 감안한다면 이는 루크(rook) 방식의 공간가중 행렬 적용과 유사하다고 할 수 있다.

먼저 모형의 설명력은 MESF 기반의 공간 가변 계수 모형의 결정계수가 0.932로 GWR 모형의 결정계수 0.776에 비해 높다. 그림 2는 각 모형의 설명력과 예측 정확도를 도표화한 것으로 가로축은 실제 아파트 매매가격(단위: 1억 원), 세로축은 추정 가격을 나타내며 기울기가

1인 추세선을 삽입하여 비교 대상으로 삼았다. 그 결과 MESF 기반 공간 가변 계수 모형의 아파트 매매가격 추정치는 매매가 15억 원에서 25억 사이의 값에서 예측 정확도가 조금 낮기는 하지만, 상대적으로 전 가격대에서 추세선 주변에 가깝게 위치하여 정확한 예측을 보여주는 것이 확인할 수 있다. 반면, GWR의 경우 10억 이하의 매매가격에서는 대부분의 점들이 추세선 위쪽에 위치하는 과대추정이, 그리고 그 이상의 매매가격에서는 대부분의 점들이 추세선 아래쪽에 위치하는 과소추정이 발생하였음을 알 수 있다(그림 2).

다음으로 각 모형 잔차의 공간적 자기상관 분석으로, 그림 3은 각각 MESF 기반 모형과 GWR 모형의 잔차(Pearson residual)를 보여준다. MESF 기반 모형의 경우

표 3. GWR과 MESF의 공간 가변 계수 간의 상관관계 행렬

MESF						
	보육시설 접근성	공원시설 접근성	지하철역 접근성	대학 진학률	특목·자립고 진학비율	방과후 프로그램 수
노후 연수	0.037	-0.029	-0.018	-0.005	-0.075**	-0.052
보육시설 접근성		-0.152***	-0.040	-0.076**	0.014	-0.087**
공원시설 접근성			0.178***	-0.088**	-0.152***	-0.100***
지하철역 접근성				-0.118***	-0.096***	-0.074**
대학 진학률					-0.122***	-0.052*
특목·자립고 진학비율						-0.038
GWR						
	보육시설 접근성	공원시설 접근성	지하철역 접근성	대학 진학률	특목·자립고 진학비율	방과후 프로그램 수
노후 연수	0.007	0.020	0.138***	-0.052*	-0.137***	0.033
보육시설 접근성		-0.424***	-0.114***	0.238***	-0.200***	-0.175***
공원시설 접근성			-0.022	-0.195***	0.156***	0.058*
지하철역 접근성				0.093***	-0.158***	0.203***
대학 진학률					0.002	0.002
특목·자립고 진학비율						0.008

* $p\text{-value} < 0.05$, ** $p\text{-value} < 0.01$, *** $p\text{-value} < 0.001$.

잔차의 전역적 모런 통계량과 그 유의수준이 0.011(0.202)로 잔차의 공간적 자기상관이 발견되지 않았다. 그에 반해 GWR 모형의 경우 모런 통계량과 그 유의수준이 0.492(<0.001)로 모형 잔차의 매우 강한 양의 공간적 자기상관을 확인하였다. 특히 상대적으로 높은 아파트 매매가격을 보이는 강남구와 그 주변에서는 양의 잔차의 군집이, 그리고 낮은 아파트 매매가격을 강서구, 도봉구, 강서구에서는 음의 잔차의 군집이 관찰된다. 이를 앞의 모형의 정확도 비교(그림 2)와 연관지어 보면, 낮은 매매가에서의 과대추정, 높은 매매가에서의 과소 추정이 있을 뿐 아니라 이들이 서로 공간적으로 인접되어 있어 GWR 모형에서는 잔차의 독립성과 등분산성 모두가 결여되었다고 할 수 있다.

마지막으로 공간 가변 계수 간의 다중공선성의 평가이다. 표 3은 각 공간 가변 계수 간의 상관관계 행렬과 그 유의수준을 보여 준다. 전제적으로 MESF 기반 모형은 공간 가변 계수들 중 7개의 변수 쌍이, 그리고 GWR 모형은 12개의 변수 쌍이 0.001보다 작은 유의수준에서의 상관관계를 보인다. 따라서 본 분석에서는 MESF 기

반의 모형에서 GWR 모형에 비해 공간 가변 계수 간의 다중공선성이 개선된 결과를 보인다고 할 수 있다. 상세히 GWR 모형의 경우 보육시설 접근성의 공간 가변 계수가 다른 모든 변수의 계수들과 유의미한 상관관계를 보이는 반면, MESF 기반 모형에서는 오직 공원의 공간 가변 계수와 유의미한 상관관계를 보인다. 하지만 공원 접근성과 특목·자립고 진학비율 계수의 경우 GWR 뿐 아니라 MESF 기반 모형에서도 다른 계수들과 유의미한 상관관계를 보이지만, 대체로 MESF 기반 모형에서 상관계수의 값이 감소하는 경향을 보인다. 이러한 계수 간의 다중공선성의 감소는 공통된 고유벡터가 각 계수의 공간적 이질성 설명하기 위해 선택되지 않고, 각 변수마다 유일한 고유벡터의 집합이 선택되었기 때문으로 설명할 수 있다(Griffith, 2008).

2. MESF 기반 공간 가변 계수를 이용한 공간적 이질성 탐색

다음은 공간 가변 계수 모형을 이용하여 각 계수의 공

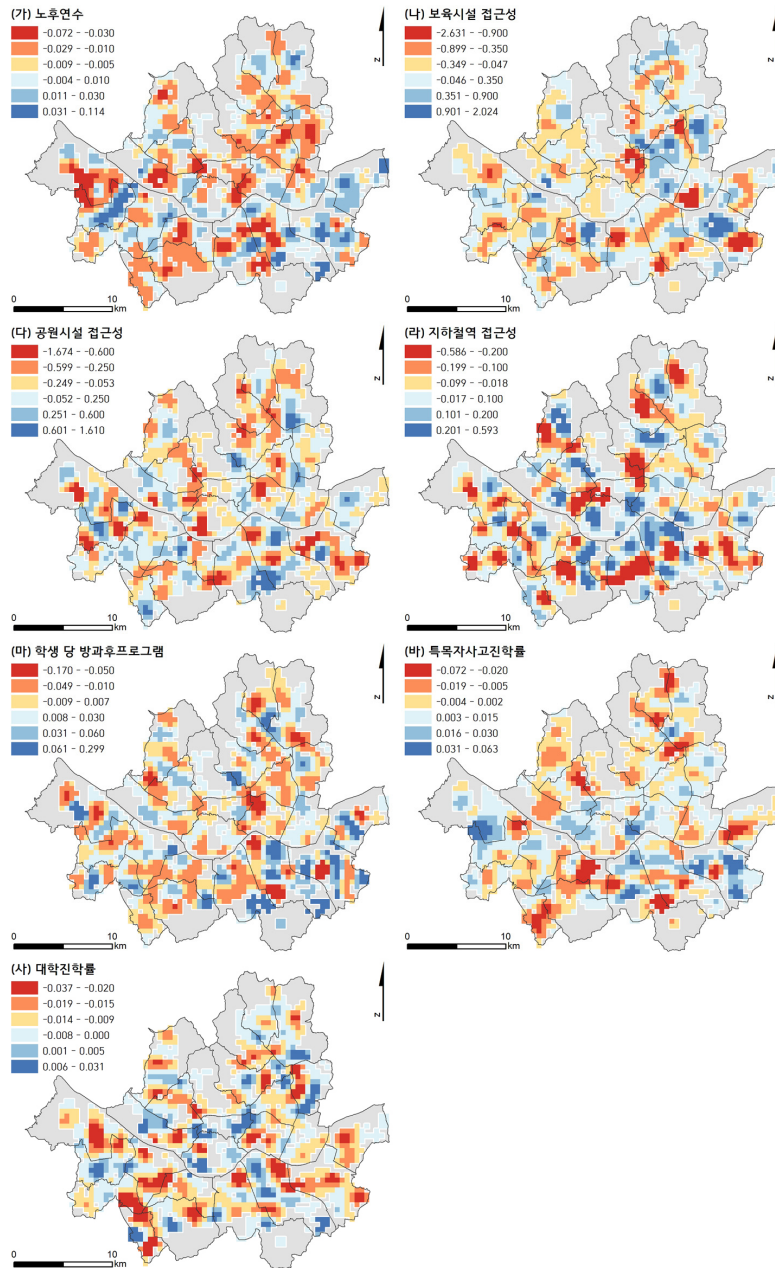


그림 4. MESF 기반 공간 가변 계수의 분포

간적 이질성을 탐색한다. 본 분석에서는 MESF 기반 모형이 GWR 모형에 비해 높은 설명력과 낮은 잔차 공간적 자기상관, 그리고 낮은 가변 계수 간의 상관관계를 보이기 때문에 다음의 공간적 이질성 탐색은 MESF 기반 모형을 이용하여 이루어진다. 더불어 식 (3)에서 보인 전

역적 계수의 값을 기준으로 각 공간 가변 계수 중 전역적 계수 값보다 큰 것은 빨간색, 그리고 작은 값은 파란색으로 지도화하여 국지적 변화를 쉽게 탐색할 수 있도록 하였다(그림 4).

먼저 아파트 노후 연수는 아파트 매매가격과 전역적으

로 유의미한 음의 상관관계를 보이거나(표 2), 국지적으로 그 계수는 $-0.072 \sim -0.114$ 로 다양하게 나타난다(그림 4의 (가)). 특히 강서구에서는 매우 강한 음의 상관관계를 보여 해당 지역에서 아파트의 노후 연수는 매매가격을 결정하는 중요한 요소로 작용할 수 있음을 확인 가능하다. 반면 강남구, 서초구, 양천구의 경우 노후화된 아파트가 많이 위치하고 있음에도 불구하고(그림 1의 (나)) 아파트 매매가격은 높아 오히려 아파트의 노후 연수와 매매가격이 양의 상관관계를 보이는 곳 또한 존재한다. 이에 반해 또 다른 노후화된 아파트가 많이 존재하는 노원구의 경우에는 그 계수의 절댓값이 앞의 구들에 비해 대체로 크지 않기 때문에 아파트의 노후 연수가 아파트 매매가격에 미치는 영향은 상대적으로 작은 것으로 판단된다.

다음은 보육시설과 공원시설의 접근성이다(그림 4의 (나), (다)). 보육시설과 공원시설 접근성의 공간 가변 계수의 범위는 각각 $-2.631 \sim -2.024$ 그리고 $-1.674 \sim -1.610$ 으로 전역적으로는 아파트 매매가격과 유의미한 상관관계를 보이지 않으나(표 2) 특정 지역에서는 강한 음과 양의 상관관계가 관찰된다. 먼저, 보육시설 접근성은 특히 서초구 방배본동, 광진구 자양동 일대에서 전역적 계수에 비해 강한 음의 상관관계가 보여 해당 지역에는 다른 지역에 비하여 보육시설의 입지가 매매가격에 중요한 역할을 하는 것으로 보인다. 반면 송파구 송파1·2동과 방이1·2동에서는 오히려 보육시설의 접근성과 아파트 매매가격 사이에 양의 상관관계가 관찰된다. 다음으로 공원시설의 접근성 경우 북한산 주변을 따라 일반적으로 전역적 계수에 비해 더 작은 국지적 계수 값이, 즉 아파트 매매가격과 더 강한 음의 상관관계가 관찰된다. 이는 해당 지역에서는 공원시설에서부터 가까울수록 아파트 매매가격이 상승하는 경향을 보인다는 것을 의미한다. 그에 반해 서초구 양재1·2동의 경우 주변에 많은 도시공원이 위치하고 있으나 해당 지역은 오히려 양의 상관관계를, 즉 도시공원이 멀수록 아파트 매매가격이 증가하는 경향을 보인다.

지하철역 접근성의 공간 가변 계수는 범위는 -0.586 에서 0.593 이며 그 분포는 매우 특징적이다(그림 4의 (라)). 먼저, 지하철역 접근성의 전역적 계수 역시 아파트 매매가격과 유의미한 음의 상관관계를 보이거나(표 2), 서울 남부지역에서는 2호선 지하철을 따라서, 특히 2호선 방배역, 봉천역, 사당역, 서초역, 그리고 신림역 근처에서 매우 강한 음의 상관관계를 보인다. 서울 북부지역에서는

특히 4호선 미아역과 수유역, 5호선 마포역과 공덕역, 그리고 6호선 보문역과 창신역 주변에서 전역적 계수보다 더 강한 음의 상관관계를 보이는 지역이 확인된다. 따라서 해당 지역은 지하철역에서 멀어질수록 아파트 매매가격의 하락폭이 상대적으로 크다고 할 수 있다. 그 외의 지역에서는 지하철역 접근성의 영향이 아파트 매매가격에 미치는 영향이 상대적으로 작다. 심지어 7호선 논현역, 학동역, 강남구청역, 그리고 청담역 주변에서는 지하철역 접근성과 아파트 매매가격 간의 양의 상관관계가 관찰되어 지하철역에서 멀수록 오히려 아파트 매매가격이 상승하는 경향을 보이기도 한다.

다음으로는 교육요인의 아파트 매매가격에 대한 공간적 이질성 탐색이다(그림 4의 (마), (바), (사)). 먼저, 학생당 방과후 프로그램은 전역적으로 아파트 매매가격에 유의미한 양의 상관관계를(표 2), 국지적으로는 $-0.170 \sim 0.299$ 사이의 공간 가변 계수의 범위를 보인다(그림 4의 (마)). 특히 강남구 북서부의 봉은, 삼성, 언북, 청담초등학교와 송파구 서부의 삼전, 잠전, 버들초등학교 지역에서는 전역적 계수 보다 더욱 강한 양의 상관관계가 관측되어, 이 지역에서는 초등학교의 교육요인이 아파트 매매가격에 더욱 강하게 영향을 미치리라 예상된다. 반면 관악구, 동작구, 양천구, 동대문구, 성동구 일대에서는 대체로 아파트 매매가격과 초등학교 방과후 프로그램 수가 음의 관계를 보이며, 특히 성동구 북부의 무학, 송신, 신당초등학교 부근에서는 강한 음의 공간 가변 계수의 군집이 관찰된다.

다음으로 중학교의 특목·자립고 진학비율은 아파트 매매가격과 전역적으로 유의미한 양의 상관관계를 보이며(표 2), 공간 가변 계수의 범위는 -0.072 에서 0.063 으로 역시 다양하게 나타난다(그림 4의 (바)). 하지만 공간 가변 계수 공간적 분포는 매우 특징적인데, 특히 상대적으로 좋은 학군이라고 알려져 있는 강남구와 서초구, 그리고 송파구 일대에서 전역적 계수 값보다 더욱 강한 양의 상관관계를 보인다. 이는 해당 지역에서 특목·자립고 진학비율의 증가가 아파트 매매가격 증가에 상대적으로 더욱 강한 영향을 미친다는 것을 의미한다. 특징적으로 양천구 북서부의 신원, 신화, 양서중학교 인근은 상대적으로 낮은 특목·자립고 진학비율을, 반면 서초구 북부의 경원, 신동, 원촌중학교 일대는 상대적으로 높은 비율을 보이지만 두 지역 모두 특히 강한 양의 상관관계가 관측된다. 이에 반해, 관악구, 동작구, 금천구에서는

일반적으로 특목·자립고 진학비율이 아파트 매매가격에 오히려 음의 상관관계를 보이는 것으로 관찰된다.

마지막으로 고등학교의 대학 진학률과 아파트 매매가격 간에는 전역적으로 유의미한 음의 상관관계가 확인되었고(표 2), 국지적으로는 -0.037에서 0.031 사이의 공간 가변 계수 범위를 보였다(그림 4의 (사)). 특히 강남구와 서초구를 포함하는 강남고등학교 군은 다른 고등학교 군에 비하여 매우 낮은 대학 진학률을 보이며(그림 1의 (아)), 해당학교 군, 특히 강남구 동부에 위치한 경기, 휘문고등학교, 그리고 서초구 북부의 반포고등학교 부근에서는 아파트 매매가격과 대학 진학률 간의 매우 강한 음의 상관관계가 확인된다. 또한 구로구, 금천구, 영등포구를 포함한 남부고등학교 군, 그리고 강동송파 학교 군에서도 전역적 계수보다 더욱 강한 음의 공간 가변 계수의 군집이 관찰된다. 반면 용산구, 중구, 종로구를 포함한 중부학교 군에서는 대부분의 지역이 전역적 계수보다 큰 공간 가변 계수, 특정지역에서는 양의 공간 가변 계수들이 관찰되어 해당지역에서는 대학 진학률이 아파트 매매가와 정적인 영향을 미친다고 할 수 있다. 같은 강남구라고 할지라도 강남구 서부는 진선여자고등학교와 은광여자고등학교는 상대적으로 높은 대학진학률(그림 1의 (아))과 신논현역과 언주역 주변의 상대적으로 낮은 아파트 매매가(그림 1의 (가))로 인하여 전역적 계수보다 큰 공간 가변 계수가 관찰된다.

IV. 요약 및 결론

아파트 매매가격 결정요인의 전역적 수준의 영향력이 각 하위지역에 동일하게 적용되지 않기 때문에 각 요인들의 공간적 이질성 탐색은 아파트 매매가격 모형의 설명력과 정확도를 향상 시킬 수 있는 하나의 방법이다. 본 연구에서는 전역적 회귀 모형의 공간적 이질성 탐색을 가능케 하는 공간 가변 계수 모형 중 하나인 MESF 기반 모형을 서울시 아파트 매매가격의 결정요인을 탐색하기 위하여 적용하고 그 결과를 GWR, 그리고 전역적 MESF 모형의 결과와 비교하였다. 상세히 전역적 회귀 모형과 공간 가변 계수 모형 모두 서울시 아파트의 매매가격을 종속변수, 그리고 아파트 크기를 오피셋으로 하는 음이항 모형에 적용하였고, 독립변수로는 아파트의 노후 연수, 보육시설의 접근성, 공원시설의 접근성, 지하

철역 접근성, 초등학교 학생 100명 당 방과후 프로그램 수, 중학교의 특수목적고 및 자립형 고교의 진학률, 그리고 고등학교의 대학 진학률을 사용하였다.

본 연구의 결과를 요약하면 다음과 같다. 먼저, MESF와 GWR 기반 공간 가변 계수 모형을 비교한 결과, MESF의 설명력(pseudo- $R^2=0.932$)이 GWR의 설명력(pseudo- $R^2=0.776$) 보다 높았다. 특히 MESF 기반 모형의 전 가격대에서 고르게 정확한 아파트 매매가격 추정을 한 반면 GWR은 낮은 가격대에서는 과대추정, 그리고 높은 가격대에서는 과소추정을 하였다. 모형 잔차의 공간적 자기상관의 유무를 고려하였을 경우에도 MESF 기반 모형에서는 잔차의 공간적 자기상관이 발견되지 않은 반면, GWR 모형의 경우 매우 강한 양의 공간적 자기상관을 확인되어 잔차의 독립성이 결여된 결과를 보였다. 마지막으로 공간 가변 계수 간의 다중공선성의 평가에서도 MESF 기반 모형에서 GWR에 비해 더 작은 수의 유의미한 상관관계를 보이는 변수 쌍이 관찰되었고, 일반적으로도 MESF 기반 모형에서 상관계수 값이 감소하는 경향을 보였다. 결론적으로 MESF 기반 모형이 모형의 설명력, 잔차의 공간적 자기상관 유무, 그리고 공간 가변 계수 간의 다중공선성 유무를 기준으로 공간적 이질성 탐색에 있어서 MESF 기반 모형이 GWR 보다 우월함을 재확인하였다(Helbich and Griffith, 2016).

다음으로 MESF 기반 공간 가변 계수를 이용한 서울시 아파트 매매가격 결정요인의 공간적 이질성 탐색 결과의 요약이다. 먼저 아파트 노후 연수는 전역적으로 유의미한 음의 상관관계를 보이나, 특히 강서구에서는 더욱 강한 음의 상관관계를, 반면 강남구, 서초구 양천구의 경우 양의 상관관계를 관찰하였다. 다음으로 보육시설과 공원시설의 접근성 모두 전역적으로는 유의미하지 않은 음의 상관관계를 보인다. 하지만 보육시설 접근성의 경우 서초구 방배본동, 광진구 자양동 일대에서 강한 음의 상관관계, 그리고 송파구 송파1·2동과 방이1·2동에서는 강한 양의 상관관계가 관찰되었다. 공원시설의 접근성 경우 북한산 주변에서 대체로 전역적 계수에 비해 강한 음의 상관관계, 반면 서초구 양재1·2동의 주변에는 양의 상관관계를 보였다. 지하철역 접근성은 전역적으로는 유의미한 음의 상관관계를 보였으며, 특히 서울 남부지역에서는 지하철 2호선을 따라서 매우 강한 음의 상관관계를 보였다. 그에 반해 강남과 서초를 지나는 지하철 7호선에서는 강한 양의 상관관계가 관찰되었다. 마지

막으로 교육요인 중 학생 당 방과후 프로그램은 전역적으로는 유의미한 양의 상관관계를, 국지적으로는 강남구 북서부에서 더욱 강한 양의 상관관계, 반면 관악구, 동작구, 양천구, 동대문구, 성동구 일대에서는 대체로 음의 계수 값을 보였다. 중학교의 특목·자립고 진학비율은 역시 전역적으로 유의미한 양의 상관관계를 보였고 특히 강남구와 서초구 그리고 송파구 일대에서 전역적 값보다 더욱 강한 양의 상관관계를 보였다. 이에 반해, 관악구, 동작구, 금천구에서는 음의 상관관계가 관찰되었다. 마지막으로 고등학교의 대학 진학률은 기존의 연구와는 다르게 전역적으로 유의미한 음의 상관관계가 확인되었고, 특히 강남구 동부와 서초구 북부에서 매우 강한 음의 상관관계가 확인되었다. 이와 같이 전역적 계수를 기반으로 아파트 매매가격에 전 연구지역에 걸쳐 동일한 영향을 미칠 것처럼 보이던 각 결정요인들이, 국지적으로 그 결정요인들의 계수 값이 매우 큰 차이를 보인다. 따라서 결정요인들의 공간적 이질성 탐색이 주택 가격 모형의 설명력 향상을 위하여 필수적임을 시사한다.

본 연구는 다음과 같은 한계를 가진다. 첫째, 아파트의 매매가격과 모든 결정요인과 관련하여, 각 변수들을 500m×500m 크기의 격자 단위로 요약하여 이용하였기 때문에 그 내부의 공간적 이질성을 파악하지 못하였다. 또한 주택의 주거환경 및 편의성 고려를 위하여 공원시설, 보육시설, 그리고 지하철역의 접근성 변수를 사용하였지만 이 접근성 변수들은 오직 거리만을 고려한 것이다. 따라서 공원시설의 면적이나, 보육시설의 선생님 수와 같은 시설의 공급이나, 해당 지역 내의 잠재적 사용자와 같은 수요의 측면을 고려하여 회귀 모형의 현실성을 향상시킬 수 있다. 교육요인을 위한 변수를 추정함에 있어서도 각 학군의 평균을 이용하기 보다는 각 학교에서 가장 가까운 직선거리를 보이는 격자의 중심에 각 학교의 정보를 할당하여 모형 추정에 이용하였기 때문에 실제 주변 교육 환경 모형화에서의 오류가 포함될 가능성이 있다. 더불어 단순 전체 대학의 진학률보다는 대학의 순위를 고려한 진학률 변수를 사용한다면 모형의 설명력을 더욱 향상시킬 수 있으리라 예상된다. 둘째, 본 분석에서는 한강의 교량으로 인한 단절된 격자 등을 수정한 후 인접성을 기반으로 이항 공간 가중 행렬을 생성한 뒤 이를 MESF 기반 모형에 사용하였다. 이를 인접된 지역 간의 차이를 나타내거나 GWR과의 연관성을 증대를 위하여 거리 기반 공간 가중 행렬로 수정이 가능하다

(Borcard and Legendre, 2002). 하지만 전역적 MESF 모형에서 후자의 결과와 전자의 결과 간의 밀접한 관련성이 확인됨으로 미루어 볼 때(Griffith and Layne, 1997) MESF 기반 국지적 모형에서도 유사한 결과를 예상해 볼 수 있다. 마지막으로 MESF는 모형 추정을 위하여 고유벡터 후보군 중 모형 설명에 적합한 고유벡터를 선택하기 때문에, 분석에 이용되는 표본의 수가 많을수록 MESF가 GWR에 비하여 일반적으로 모형 추정에 더 많은 계산 시간이 요구된다. 본 분석에서는 상대적으로 계산에 오랜 시간이 필요한 단계적 회귀 분석 방법을 이용하였지만, 유전 알고리즘(genetic algorithm)(Helbich and Griffith, 2016)이나 축소추정법(Seya *et al.*, 2015)을 이용하면 그 계산 시간을 단축할 수 있다. 또한 MESF 기반 모형과 LASSO나 능형 회귀 모형과 결합된 GWR 간의 다중공선성의 해소 정도를 직접적으로 비교해보는 것 또한 향후 연구의 주제가 될 수 있다.

註

- 1) SAAR ver.1.0.7을 이용하였다.
- 2) ESF가 적용되지 않은 포아송과 음이항 회귀 모형의 편차 통계는 각각 7488.87과 1.02이다. 일반 음이항 회귀 모형에서의 추정된 산포 패러미터(dispersion parameter)는 0.117이다.
- 3) 국토교통부 실거래가 공개시스템.
- 4) 국토정보플랫폼 생활기반시설 접근성 지표.
- 5) 소공원, 어린이공원, 근린공원 등의 생활권 공원과 역사공원 문화공원, 체육공원 그 밖의 조례에 의한 주제공원을 모두 포함한다.
- 6) 국가교통데이터베이스.
- 7) 학교알리미.
- 8) 전역적 모런 통계량과 그 유의수준은 0.6470 (<0.01)이다.
- 9) spgwr: Geographically Weighted Regression.
- 10) 잔차의 공간적 자기상관을 제거하기 위하여 고유벡터의 후보군을 확장할 필요성이 있다. 하지만 후보군을 증가시키면 공간 가변 계수 추정 시 더 많은 컴퓨팅 자원이 요구되므로 그 자원 증가를 막고 공간 가변 계수 모형과의 전역 모형과의 직접 비교를 위해 현재의 고유벡터 후보군을 유지하였다.

참고문헌

- 강수진·서원석, 2016, “지하철 노선 및 역사특성이 아파트 매매가격에 미치는 영향 분석” 국토연구, 89, 149-162.
- 강창덕, 2010, “GWR 접근법을 활용한 부동산 감정 평가 모형 연구: 서울시 아파트를 사례로” 부동산연구, 20(2), 107-132.
- 구형모, 2018, “대학 진학 이동과 대학 졸업자 취업 이동의 공간적 특성 탐색” 한국도시지리학회지, 21(3), 109-122.
- 김강영, 2016, “주택시장지역 구획을 위한 공간 최적화 모형” 한국지도학회지, 16(1), 53-65.
- 김강영, 2017, “아파트 매매가격 변동의 시공간적 역동성 탐색” 한국지도학회지, 17(2), 45-62.
- 김경근, 2019, “변화의 시대에 교육은 미래 세대의 삶을 적절히 준비하고 있는가?” 한국사회학회 심포지움 논문집, 73-103.
- 김경민·이양원, 2007, “사교육시장 및 교육성도가 아파트가격에 미치는 영향: 2004년 이후 아파트가격 상승기를 중심으로” 국토연구, 55, 239-252.
- 김민성·박세운, 2014, “공간회귀모형을 이용한 아파트 가격 결정요인에 관한 연구” 한국산업경제학회 정기학술발표대회 초록집, 113-132.
- 김소연·김영호, 2013, “주거지 인문환경의 공간 속성을 고려한 주택 가격 결정 모형: 서울시 아파트를 대상으로” 한국지도학회지, 13(3), 41-56.
- 김익희, 2005, “GWR을 이용한 지가설명함수의 공간적 이질성(Spatial heterogeneity) 탐색: 서울시 서초구 강남구 송파구를 사례로” 서울대학교 석사학위논문.
- 김정희, 2018, “공간-통계 융합 자료를 이용한 공동주택의 가격결정모형 비교 분석” 한국사진지리학회지, 28(1), 13-26.
- 김혜영·전철민, 2012, “공간구문론 및 지리적 가중회귀 기법을 이용한 지가분석” 한국지리정보학회지, 15(2), 35-45.
- 남현권·서원석, 2017, “지리가중회귀모형을 이용한 아파트 점유형태별 전철역의 접근성 가치 비교분석” 서울도시연구, 18(2), 65-79.
- 박운선·임병준, 2010, “헤도닉 가격모형을 활용한 아파트 가격결정요인 분석: 서울시 및 부산시를 중심으로” 대한부동산학회지, 28(2), 245-271.
- 이상일·조대현·이민파, 2015, “일변량 공간연관성통계량에 대한 비교연구 (I): 전역적 S 통계량을 중심으로” 한국지리학회지, 4(2), 329-345.
- 이상일·조대현·이민파, 2016, “일변량 공간연관성통계량에 대한 비교연구 (II): 국지적 S 통계량을 중심으로” 한국지리학회지, 5(3), 375-396.
- 임준홍·홍성효, 2015, “교육성취의 아파트가격에 대한 영향: 서울시 사례분석” 부동산학연구, 21(1), 75-89.
- 전경구, 2012, “교육환경이 아파트 가격에 미치는 상호작용 효과분석” 한국지역개발학회지, 24(3), 1-24.
- 전창우·조대현·주뢰, 2018, “지리가중능형회귀(GWR)를 이용한 미세먼지(PM10)의 공간적 이질성 탐색” 한국지도학회지, 18(3), 91-104.
- 정수연, 2006, “교육요인이 서울아파트가격에 미치는 영향에 관한 연구” 국토계획, 41(2), 153-166.
- 조대현, 2018, “주택가격과 인구이동 간의 연관성에 관한 공간 분석: 수도권의 전세가격을 중심으로” 한국지리학회지, 7(3), 449-462.
- 진찬우·이건학, 2014, “GWL을 적용한 공간 헤도닉 모델링” 대한지리학회지, 49(6), 917-934.
- 홍하연·이주형, 2015, “아파트 가격에 영향을 미치는 요인의 시공간적 영향력 변화 연구: 서울시 25개 구를 대상으로” 서울도시연구, 16(2), 87-108.
- Borcard, D. and Legendre, P., 2002, All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices, *Ecological Modelling*, 153(1-2), 51-68.
- Brent, R., 1973, *Algorithms for Minimization without Derivatives*, Englewood Cliffs, NJ: Prentice-Hall.
- Casetti, E., 1972, Generating models by the expansion method: Applications to geographical research, *Geographical Analysis*, 41(1), 81-91.
- Casetti, E., 1997, The Expansion method, mathematical modeling, and spatial econometrics, *International Regional Science Review*, 20(1-2), 9-33.
- Chun, Y. and Griffith, D.A., 2013, *Spatial Statistics and Geostatistics: Theory and Applications for Geographic Information Science and Technology*, London: Sage Publications.
- Fotheringham, A.S., Brunson, C., and Charlton, M., 2002, *Geographically Weighted Regression: The Analysis*

- of *Spatially Varying Relationships*, Chichester, England: John Wiley & Sons.
- Griffith, D.A., 2002, A spatial filtering specification for the auto-Poisson model, *Statistics and Probability Letters*, 58(3), 245-251.
- Griffith, D.A., 2003, *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, Berlin: Springer.
- Griffith, D.A., 2008, Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR), *Environment and Planning A*, 40(11), 2751-2769.
- Griffith, D.A. and Haining, R., 2006, Beyond mule kicks: The Poisson distribution in geographical analysis, *Geographical Analysis*, 38(2), 123-139.
- Griffith, D.A. and Layne, L., 1997, Uncovering relationships between geo-statistical and spatial autoregressive models, in American Statistical Association, ed., *The 1996 Proceedings of the Section on Statistics and the Environment*, Alexandria, VA: American Statistical Association, 91-96.
- Helbich, M. and Griffith, D.A., 2016, Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches, *Computers, Environment and Urban Systems*, 57, 1-11.
- Koo, H., Chun, Y., and Griffith, D.A., 2018, Integrating spatial data analysis functionalities in a GIS environment: Spatial analysis using ArcGIS engine and R (SAAR), *Transactions in GIS*, 22(3), 721-736.
- Lin, G. and Zhang, T., 2007, Loglinear residual tests of Moran's I autocorrelation and their applications to Kentucky breast cancer data, *Geographical Analysis*, 39(3), 293-310.
- Redfearn, C.L., 2009, How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets, *Regional Science and Urban Economics*, 39(3), 297-306.
- Seya, H., Murakami, D., Tsutsumi, M., and Yamagata, Y., 2015, Application of LASSO to the eigenvector selection problem in eigenvector-based spatial filtering, *Geographical Analysis*, 47(3), 284-299.
- Tiefelsdorf, M. and Boots, B., 1995, The exact distribution of Moran's I, *Environment and Planning A*, 27(6), 985-999.
- Wheeler, D., 2009, Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso, *Environment and Planning A*, 41(3), 722-742.
- Wheeler, D. and Tiefelsdorf, M., 2005, Multicollinearity and correlation among local regression coefficients in geographically weighted regression, *Journal of Geographical Systems*, 7(2), 161-187.
- 국가교통데이터베이스, <https://www.ktdb.go.kr/www/index.do>
- 국토교통부 실거래가 공개시스템, <http://rt.molit.go.kr>
- 국토정보플랫폼 생활기반시설 접근성 지표, <http://map.ngii.go.kr/ms/life/lifeInfrastructureMap.do>
- 서울교통공사, “서울시 역코드로 지하철역 위치 조회”, <https://data.seoul.go.kr>
- 학교알리미, <https://www.schoolinfo.go.kr>
- SAAR(ver.1.0.7), <https://thesaar.github.io>
- spgwr: Geographically Weighted Regression, <https://cran.r-project.org/web/packages/spgwr/index.html>
- 교신 : 구형모, 210046, 중국 장쑤성 난징시 치샤구 원위 안로 1, 난징사범대학교 지리과학학원(이메일: hmo.koo@gmail.com)
- Correspondence : Hyeongmo Koo, 210046, 1 Wenyuan Road, Qixia District, Nanjing, Jiangsu, China, School of Geography, Nanjing Normal University (Email: hmo.koo@gmail.com)
- 투 고 일: 2019년 6월 27일
심사완료일: 2019년 7월 22일
투고확정일: 2019년 7월 31일